Answers

Chapter 2 Review Problems

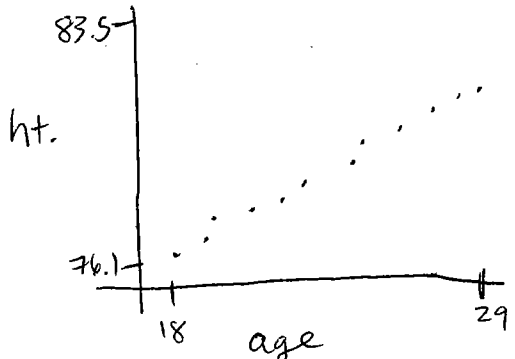
Use the data below for questions 1 through
Below is data concerning the mean height of Kalama children. A scientist wanted to look at the effect that age had on the mean height of the children.

Age (months)	18	19	20	21	22	23	24	25	26	27	28	29
Height (cm)	76.1	77	78.1	78.2	78.8	79.7	79.9	81.1	81.2	81.8	82.8	83.5

1. Determine the explanatory and response variables

expl = AGE resp = HEIGHT

2. Create a scatterplot of the data. Be sure to label the axes. DESCRIBE the plot.



- positive
- linear
- strong

3. Find the LSR line and the correlation coefficient

$\hat{y} = 64.93 + 0.635x$ $r = 0.994$

4. What proportion of the variability in the height of Kalama children is explained by the variability in their age?

$r^2 = 0.9888$

5. Interpret the slope of the LSR line in a complete sentence

$\frac{\Delta y}{\Delta x} = \frac{0.635 \text{ cms}}{1 \text{ month}}$ = For every one month a kid ages, it grows 0.635cms on average.

6. Predict the mean height of a child that is 42 months old (show work!). Is this prediction accurate? Why or why not?

- no extrapolation

$\hat{y} = 64.93 + 0.635(42)$

$\hat{y} = 91.5969$ cms

7. Predict the mean height for a child who is 24 ~~months old~~ (show work!)

$\hat{y} = 64.93 + 0.635(24)$ $\hat{y} = 80.167$

8. Find the error of your predicted value for a child who is 24 months old

$79.9 - 80.167 = -0.2675$

9. Was your prediction an overestimate or an underestimate?

Over estimate

10. How old is a child expected to be if they are 100cm long? (show work!)

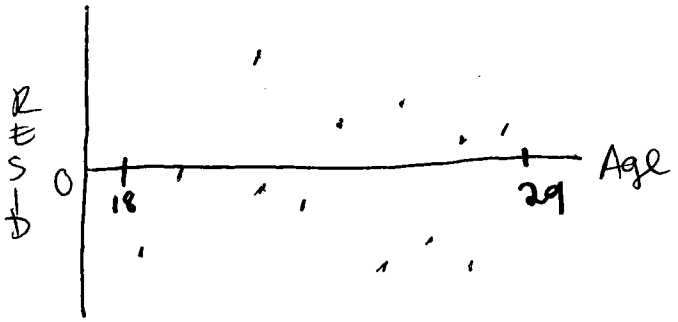
$100 = 64.93 + 0.635x$

$x = 55.23$ months

11. What is the sum of the residuals for this data?

0 (6×10^{-12})

12. Create a residual plot below. Describe the plot.



13. What does the plot tell you about your linear model? Explain BRIEFLY.

Scattered = good model for data.

14. What conclusions can be made from the previous questions? Does the age of a child CAUSE the height of the child? Why or why not?

Somewhat → high r, r^2
plausible
Scattered resid plot.

15. Below is a Minitab statistical analysis. The data is looking at clothes salespersons and examining the effect that the number of minutes spent with a customer has on the total dollar amount that the customer buys. In other words, if a salesperson spends more time with a customer, does the customer buy more clothing (increasing the commission of the salesperson)?

Predictor	**Coeff**	**s.e.**	**T**	**P**
Constant	-1.731	2.4065	-0.876	0.4561
Minutes	0.5679	0.00456	6.6898	1.2358

S = 1.3425 R-Sq = 0.7896 R-Sq (adj) = 0.7748

(a) What is the equation of the LSR line?

$\hat{y} = -1.731 + 0.5679x$

(b) What is the value of the correlation coefficient?

$r = \sqrt{0.7896} \neq 0.88859$

(c) What does the correlation tell you about the relationship of your two variables?

- positive
- strong
- linear

(d) Interpret the slope in the context of the problem

$\frac{\Delta y}{\Delta x} = \frac{\$0.5679}{1 \text{ min}}$ = For every 1 min spent w/ customer, the amount spent increases by $0.57.

(e) What is the coefficient of determination? Interpret this value in context of the problem.

$r^2 = 0.7896$

78.96% of the variation in amount spent on clothes is due to amount of time spent w/ customer

16. Given the following data about variables x and y calculate by hand (using AP formulas) the LSR line. Show all work! Write the line in the form y = a + bx .

	X	Y
Mean	45.6	37.2
St. Dev	3.2	2.1

r = 0.765

$b = r\frac{S_y}{S_x} = 0.765\left(\frac{2.1}{3.2}\right) = 0.502$

$\hat{y} = 14.31 + 0.502x$

$a = \bar{y} - b\bar{x} = 37.2 - (0.502)(45.6) = 14.31$

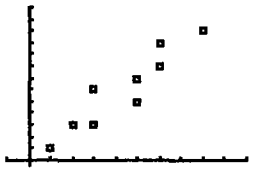
17. What does a residual plot tell us? What do we look for in a residual plot?

the fit of the linear model — scattered!

18. What type of relationship does r represent?

LINEAR!

19. For the graph below, what would be the closest approximation to the correlation coefficient?



(a) 0.2
(b) 0.88
(c) –0.9
(d) –0.2
(e) 0
(f) 0.5

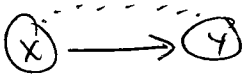
20. What is the difference between outliers and influential observations?

Outliers
- extreme in y-direction
- large residuals
- outside overall pattern

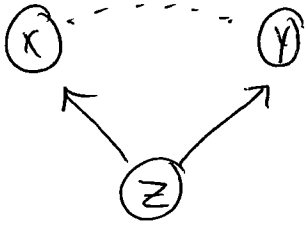
Influential
- extreme in x-dir.
- significantly changes LSR and r if removed
- pulls LSR toward it
- no other pts. near it.

21. Explain common response, confounding and causation in your own words, and draw diagrams for each

Causation

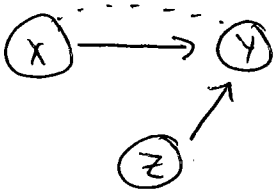


Common Resp.



Ex: TV vs. Life Exp.

Confounding



Ex: wine vs. heartdis.

22. Explain Simpson's Paradox in your own words

Make one conclusion based on overall data, but make a different conclusion when another variable is added

23. Explain Independence (in categorical distributions) in your own words

Marginal distributions match conditional distributions for each variable.

24. Describe the following plots:

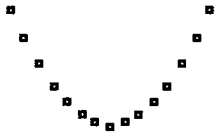
(a)

- negative
- curved
- moderate

(b)

- 2 clusters
- pos. assoc.
- weak linear

(c)



- curved
- strong
- neg, then pos.

AP Statistics

Section 2.6 – Categorical Data (Example)

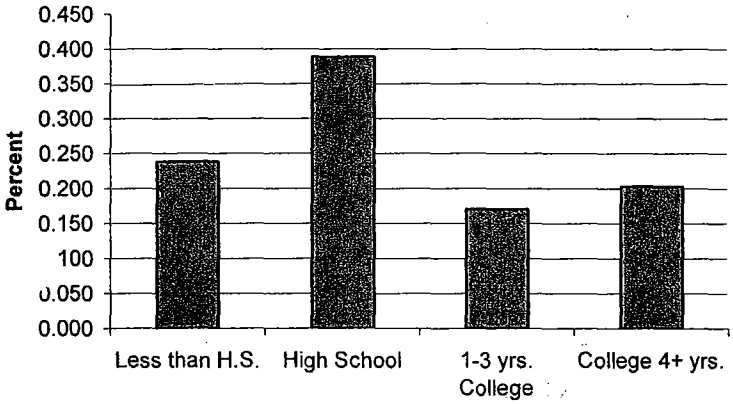
YEARS OF SCHOOL COMPLETED vs. AGE, 1998

EDUCATION	25-34	35-44	45-54	55-64	65+	TOTAL
Less than H.S.	5,836	4,841	5,230	7,024	13,183	36,114
High School	17,889	13,200	9,860	8,580	9,412	58,921
1-3 yrs. College	9,069	7,309	3,698	2,793	2,915	25,784
College 4+ yrs.	10,174	9,332	5,008	3,246	3,018	30,781
TOTAL	42,968	34,682	23,796	21,643	28,528	151,616

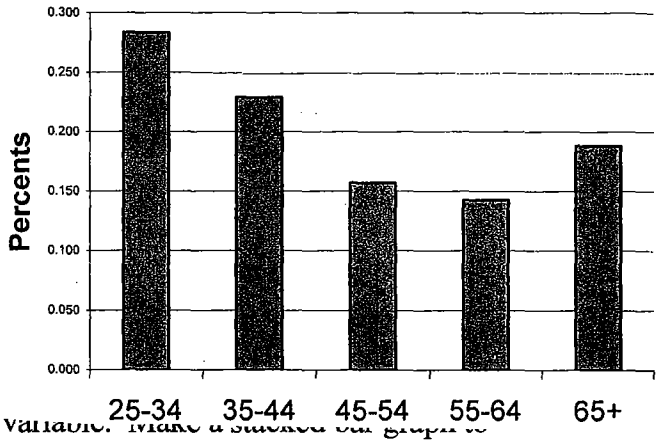
THOUSANDS OF PEOPLE

1. Calculate the **marginal distributions** (one for education, and one for age, show the numbers). Create graphs to display each distribution.

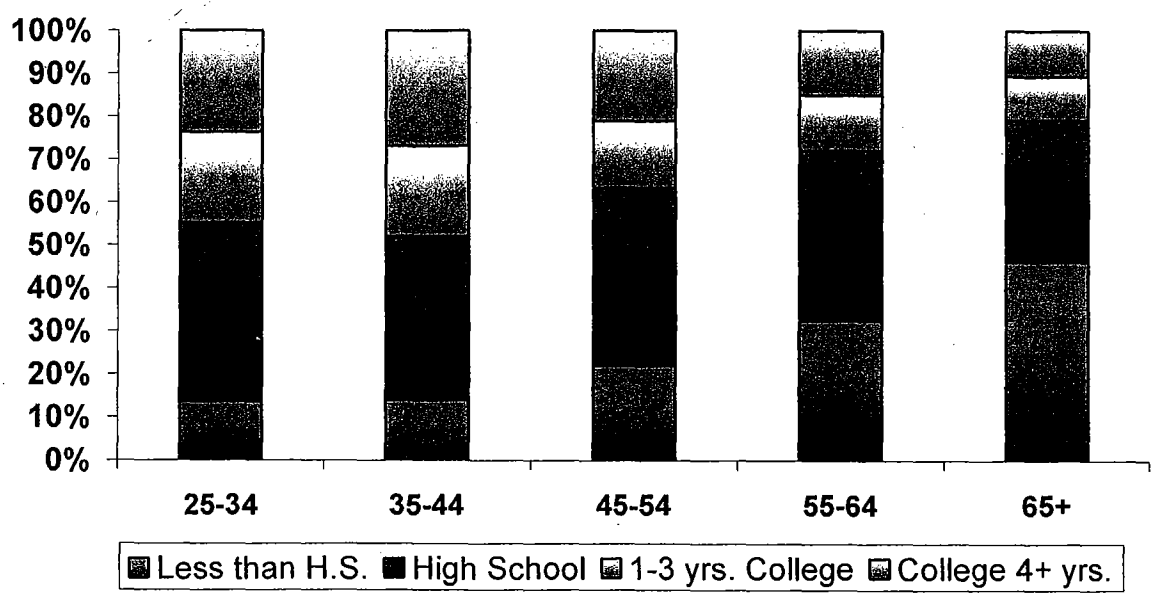
Marginal Distribution of Education



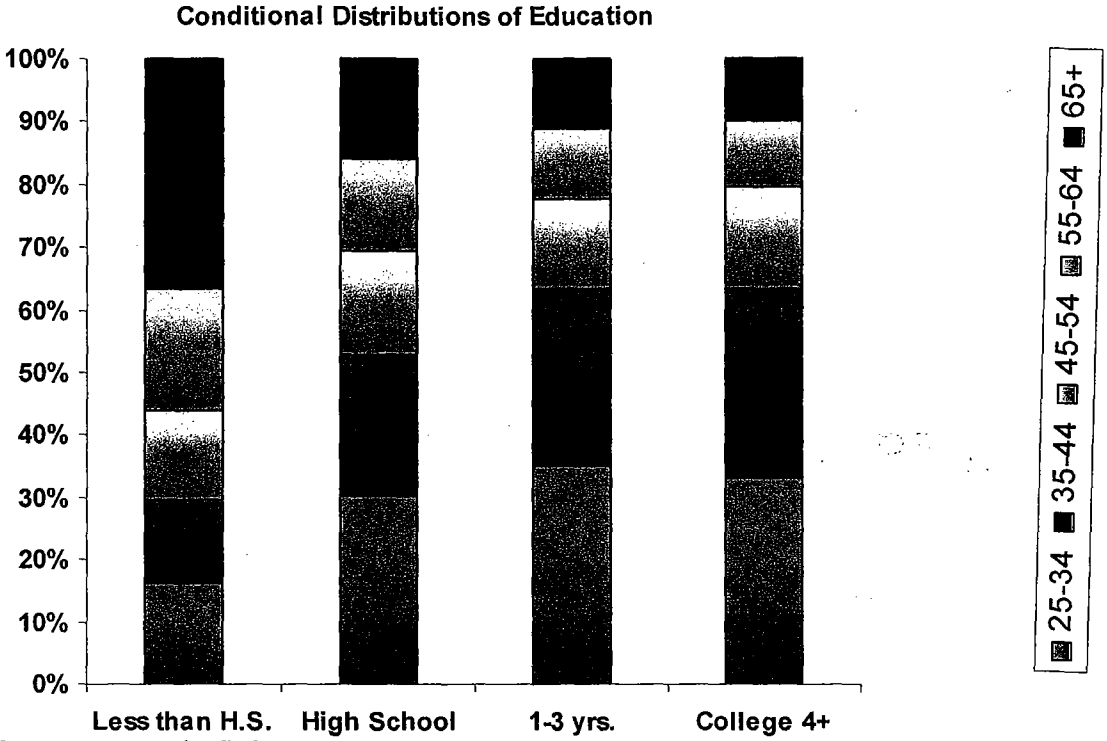
Marginal Distribution of Ages



2. Calculate the **conditional distribution** for the age variable. Make a stacked bar graph to display the information. Describe and make conclusions from the graph.



3. Calculate the **conditional distribution** for the education variable. Create a stacked bar graph. Describe the distribution and make conclusions from it.



a. Of those who only finished high school, what percent were 25 – 34?

17889/58921 = 0.3036

b. Of those people older than 64, what percent have at least been to college?

(2915+3018)/28528 = 0.208

c. What percent of people have only graduated high school?

58921/151616 = 0.3886

d. What percent of people are less than 45?

(42968 + 34682)/151616 = 0.512

e. What percent of people finished only high school OR were 25-34?

(58921 + 5836 + 9069 + 10174)/151616 = 0.554

f. What percent of people are older than 64 and have at least been to college?

(2915 + 3018)/151616 = 0.039